Accepted Manuscript

Title: Reproducibility of the Balance Evaluation Systems Test (BESTest) and the Mini-BESTest in school-aged children

Authors: R. Dewar, A.P. Claus, K. Tucker, R. Ware, L.M. Johnston

PII: S0966-6362(17)30124-8
DOI: http://dx.doi.org/doi:10.1016/j.gaitpost.2017.04.010
Reference: GAIPOS 5383

To appear in: *Gait & Posture*

Received date: 26-10-2016
Revised date: 3-4-2017
Accepted date: 5-4-2017

Running Head: Reproducibility of BESTest in children

Reproducibility of the Balance Evaluation Systems Test (BESTest) and the Mini-BESTest in school-aged children

R Dewar BPhty (Hons)[1], A P Claus PhD[1], K Tucker PhD[2], R Ware PhD[3,4], L M Johnston PhD[1]

[1] *The University of Queensland, School of Health and Rehabilitation Sciences, Brisbane, Australia,* [2] *The University of Queensland, School of Biomedical Sciences, Brisbane, Australia,* [3] *Griffith University, Menzies Health Institute Queensland,* [4]*The University of Queensland, Queensland Centre for Intellectual and Developmental Disability, Brisbane, Australia.*

Correspondence: Rosalee Dewar, School of Health and Rehabilitation Sciences, The University of Queensland, Brisbane QLD 4072, Australia. Phone: +61 7 3365 2791.

Email: rosalee.sheather@uq.net.au

HIGHLIGHTS

- First to examine reproducibility of the Full-BESTest and Mini-BESTest in children
- Both tests can discriminate postural control abilities in children
- Both tests have the potential to measure change over time
- The comprehensive Full-BESTest shows marginally better reproducibility
- A modified version of the BESTest for kids is proposed (Kids-BESTest)

Abstract

This study evaluated the intra-rater, inter-rater and test-retest reproducibility of the Full-BESTest and Mini-BESTest when assessing postural control in children. Thirty-four children aged 7-17 years participated in intra-rater and inter-rater evaluation, and 22 children repeated assessment six weeks later for evaluation of test-retest reliability. Postural control was assessed using the Full Balance Evaluation Systems Test (Full-BESTest) and the short-form Mini-BESTest. Intra-rater, inter-rater and test-retest reproducibility were examined using video assessment. Test-retest reproducibility was also assessed in real-time. Reproducibility was examined by agreement and reliability statistics. Agreement was calculated using percentage of agreement, Limits of Agreement and Smallest Detectable Change. Reliability was calculated using Intra-class Correlation Coefficients. Results showed that the reliability of *Total Scores* was excellent for the Full-BESTest for all conditions (all ICCs>0.82), whereas the Mini-BESTest ranged from fair to excellent (ICC=0.56 to 0.86). Percentage of *Domain Scores* with good-excellent reliability (ICCs>0.60) was slightly higher for the Full-BESTest (66%) compared to the Mini-BESTest (59%). Smallest Detectable Change scores were good to excellent for the Full-BESTest (2% to 6%) and for the Mini-

BESTest (5% to 10%) relative to total test scores. Both the Full-BESTest and Mini-BESTest can discriminate postural control abilities within and between days in school-aged children. The Full-BESTest has slightly better reproducibility and a broader range of items, which could be the most useful version for treatment planning. We propose minor modifications are recommended to improve reproducibility for children, and indicate the modified version by the title Kids-BESTest. Future psychometric research is recommended for specific paediatric clinical populations.

Keywords: Postural Control; Children; Reproducibility; BESTest; Mini-BESTest

1. Introduction

Postural control is commonly defined as the ability to control the body's position in space for the purpose of postural orientation and postural stability [1]. Postural control depends on the integration of sensory, motor and cognitive systems [2], and deficits in postural control may result from impairments in any or all of these systems [3]. Deficits in postural control have been shown to contribute to activity limitations experienced by children with a wide range of conditions, for example: Cerebral Palsy [4], Developmental Coordination Disorder [5-7], Spina Bifida [8], Down syndrome [9], Autistic Spectrum Disorders [10], premature birth [11, 12] and sensorineural hearing loss [13]. However, each of these studies has examined only certain sub-components of postural control due to the lack of a comprehensive clinical assessment for children.

Using children with neurological disorders as an example, it can be seen that it is clinically important to measure potential deficits in all systems involved in postural control. For example, children with cerebral palsy have demonstrated deficits in anticipatory mechanisms (feedforward postural adjustments) [14-18], adaptive mechanisms (feedback postural adjustments) [3, 14, 19-21], musculoskeletal systems (muscle force and range of motion required for standing balance) [22] and sensory systems (visual and proprioceptive function required for balance) [23, 24]. This research illustrates well how one population can experience a broad range of postural control problems, and also that there is an absence of a comprehensive clinical assessment for children. Of the clinical assessments that exist to measure postural control impairments in children [25-27], none assess all systems involved in postural control [25, 28], and many have limited psychometric data [26, 28, 29]. Research on postural control deficits in multiple populations [9, 25, 27, 28, 30] flags the need to identify or develop a comprehensive assessment battery with associated normative data against which postural control function of children with disabilities can be measured.

One existing assessment worthy of consideration for children is the Balance Evaluation Systems Test (BESTest), which is a comprehensive postural control test battery developed to evaluate impairments in adults with brain injury [31]. In adults, the original, or *Full-BESTest*, has been shown to assist in identifying specific balance deficits [32] and in measuring changes in postural control before and after intervention [33]. In children, one preliminary study involving children with CP has supported the potential use of the Full-BESTest with paediatric populations [34]. However, there is a need to establish psychometric data and clinical utility for children with and without disabilities before more widespread use [26, 28].

The Full-BESTest contains 36 items that identify and classify postural control deficits across six domains or systems of postural control: *Biomechanical constraints, Stability Limits/Verticality*, *Transitions - Anticipatory Postural Adjustment, Reactive Postural Response*, *Sensory Orientation* and *Stability in Gait* [31]. Items in each domain are scored from 0 (worst performance) to 3 (best performance) to yield a *Total Score* out of 108 points. *Domain scores* can also be calculated. For situations where clinical time limits may exclude the use of the Full-BESTest (which requires 30 min), a short-form, or *Mini-BESTest*, has been developed with a subset of 14 items (which requires only 10-15 min) [35]. In the Mini-BESTest each item is scored from 0 (worst performance) to 2 (best performance) with a maximum of 28 points. A potential limitation of the Mini-BESTest is that only 4 of the 6 domains are represented, *Anticipatory Postural Adjustments*, *Reactive Postural Response*, *Sensory Orientation* and *Stability in Gait*, to focus on the construct of 'dynamic balance' [36]; However it is still worthy of consideration as a potential screening tool for postural control impairment.

This study examined the reproducibility of the Full-BESTest and the Mini-BESTest for assessing postural control in typically developing school-aged children. Reproducibility evaluates the degree to which repeated measurements provide similar results and includes two components: (i) agreement and (ii) reliability [37]. Agreement assesses how close the results of repeated measurements are, and the margins that represent real clinical change, as opposed to random measurement error [37]. Reliability assesses whether participants can be distinguished from each other on the basis of performance, despite measurement error [37]. Qualitatively, the process of performing the Full-

BESTest and Mini-BESTest was also considered to determine which of the BESTest versions were feasible for use with children.

2. Methods

2.1 Study Design and Participants

Intra-rater, inter-rater and test-retest reproducibility of the Full-BESTest and Mini-BESTest were examined with typically developing school-aged children in June-July 2015. Ethical approval was obtained from Human Research Ethics Committees of The University of Queensland (EC00179) and the Cerebral Palsy League (EC00417), in accordance with National Health and Medical Research Council's (NHMRC) guidelines.

Participants were volunteers sought from the community using flyers and newsletter advertisements. Children were eligible for inclusion if they had typical development and were aged between 7-18 years. Typical development was confirmed using the Bruininks-Oseretsky Test of Motor Proficiency, second Edition (BOT-2) short form. Participants were excluded if they: (i) achieved a percentile rank of ≤ 5% on the BOT-2, (ii) had another known medical or behavioural disorder that may impact results, or (iii) were born at < 36 weeks gestation. Individual information forms were provided to children and their guardians, along with verbal explanations of the protocol. All guardians signed consent forms and all children signed assent forms prior to participation.

2.2 Outcome measures

The Full-BESTest was administered as per the original instructions published by Horak and colleagues in 2009 [31], except for minor modifications to instructions on a few items needed for these to be understood by children (Appendix 1). The Mini-BESTest was scored exactly as per the protocol described by Franchignoni and colleagues in 2010 [35].

2.3 Procedure

Real time and video data collection was completed on Day 1 (n=34) and Day 2 (n=22) by the primary author (RD) who is a senior physiotherapist with 19 years of clinical experience in paediatrics and management of children with CP. The interval between assessments was 2-6 weeks. This enabled reproducibility evaluation from intra-rater (video, n=34) and test-retest (real time and video, n=22) perspectives. A second senior paediatric physiotherapist with 20 years of experience re-rated the Day 1 videos to enable calculation of inter-rater reproducibility (n=34). The second rater was not a study investigator and so was considered an independent examiner. To ensure assessment fidelity, both raters completed standardized online training via the BESTest website prior to performing data collection and/or extraction [38] and a video recording protocol was used for data collection (Appendix 2).

2.4 Data Analysis

Reproducibility (agreement and reliability) of Full-BESTest and Mini-BESTest ratings was examined under four conditions: (1) Day 1 intra-rater video (n=34); (2) Day 1 inter-rater video (n=34); (3) Day 1 and 2 test-retest video (n=22) and (4) Day 1 and 2 test-retest in real-time (n=22). All parameters of reproducibility listed below were calculated for the *Total Score* and all *Domain Scores* for both the Full-BESTest (7 domains) and the Mini-BESTest (4 domains).

Agreement analysis involved calculation of percentage of exact agreement (%EA), Standard Error of Measurement (SEM), Smallest Detectable Change (SDC) and Limits of Agreement (LoA). The %EA was calculated to indicate the percentage of scores that were the same between Day 1 and Day 2 assessments. A priori, the level of clinically suitable percentage agreement for the Full-BESTest was set at excellent if >90% within 4 points, good if >80% within 4 points, fair if >60% within 4 points or poor if <60% within 4 points. This agreement level was chosen to accommodate a minor and not clinically significant variation expected from this test, which has 37 items and a scoring range of 108 points. The a priori level of agreement set for the Mini-BESTest was set at excellent if >90% within 2 points, good if >80% within 2 points, fair if >60% within 2 points or poor if <60% within 2 points. This setting was lower to reflect the fewer items (n=14) and more narrow scoring range (28 points) of the Mini-BESTest. The SEM was calculated to indicate the measurement error of the BESTest and in turn this was used to calculate the SDC, which is the smallest change in score that will represent real change and not just measurement error [37]. The SDC was expressed as a percentage to facilitate comparison between domains and total scores with different ranges. For the purposes of this study, an a priori SDC of 0-5% was considered to be excellent agreement; >5-10% to be good, >10-15% to be fair and > 15% to be poor agreement. Finally, the LOA with 95% confidence interval was calculated to describe the range within which similar scores were produced by different raters or the same rater on separate occasions.

Reliability was calculated via Intra-class Correlation Coefficients (ICC) and 95% confidence intervals using analysis of variance models. The ICC values were interpreted according to recommended criteria [39] previously used in studies with children [40], whereby an ICC > 0.75 was considered to be excellent, 0.74 – 0.60 to be good, 0.59 – 0.40 to be fair and < 0.4 to be poor.

3. Results

A total of 34 children who met the inclusion / exclusion criteria attended for initial assessment (Table 1). They were aged from 7 years 10 months to 17 years 6 months (mean 10 years 10 months; 56% male). Of these, 22 children returned for the repeat assessment (Table 1).

3.1 Full-BESTest

3.1.1 Intra-rater reproducibility (video assessment)

Intra-rater reproducibility evaluation of the Full-BESTest showed excellent agreement (100% within 2 points, Table II) and excellent reliability (*Total score* ICC=0.96, 95% CI 0.93 to 0.99; *Domains* ICC = 0.80 to 0.90, Table III) for the *Total Score* and all *Domains* using video assessment. The SDC was excellent (2.2 points, 2%) indicating that children must improve by 3 points to demonstrate real change when examined by one examiner.

3.1.1 Inter-rater reproducibility (video assessment)

The Full-BESTest *Total Score* showed excellent inter-rater agreement (>90% within 4 points, Table II) and excellent reliability (ICC = 0.87, 95% CI 0.79 to 0.95) (Table III) using video-based assessment. The *Domain Scores* showed good agreement (88-100% within 2 points) and at least fair reliability for 4 domains (Table II). Two domains showed poor reliability, *Sensory Orientation and Biomechanical Constraints*, but excellent agreement (100% agreement within 2 points) due to ceiling effects that reduced ICC scores in both domains. The SDC for the Full-BESTest *Total score* was excellent (4 points, 4%) indicating that children must improve by 4 points to demonstrate real change when rated by different examiners.

3.1.3 Test-retest reproducibility (real-time and video assessment)

The Full-BESTest *Total score* showed slightly better test-retest agreement for video (64% within 2 points) compared to real-time assessment (54% within 2 points) (Table II). Reliability was excellent for both video (ICC = 0.84, 95% CI 0.72 to 0.96) and real-time assessment (ICC = 0.82, 95% CI 0.69 to 0.96) (Table III). The *Domain scores* showed excellent agreement for video (91-100% within 2 points depending on mode) and for real-time assessment (86-100% within 2 points) (Table II). Three domains showed excellent reliability for video and three showed good-excellent reliability in real-time. *Sensory Orientation* showed poor reliability, but good agreement in real-time, again due to a ceiling effect. *Reactive Postural Response* showed the lowest agreement and poor reliability in both assessment modes due to variability of performance of children in some items in this section. The SDC was excellent for video (SDC=4.9, 5%) and good for real-time assessment (6.6, 6%), indicating that when scored using video on two different occasions children must improve by 5 points on the 108-point scale to demonstrate real change.

3.2 Mini-BESTest

3.2.1 Intra-rater reproducibility (video assessment)

The Mini-BESTest *Total Score* and all *Domains* showed excellent intra-rater agreement (100% within 2 points, Table IV) and the *Total Score* also showed excellent reliability (ICC 0.86, 95% CI 0.78 to 0.95, Table V) using video assessment. The domains showed good-excellent reliability. *Sensory Orientation* achieved perfect agreement and so an ICC could not be calculated. The SDC was good (1.50, 5%) indicating that children must improve 2 points to demonstrate real change when examined by one examiner.

3.2.2 Inter-rater reproducibility (video assessment)

The Mini-BESTest *Total Score* showed good inter-rater agreement (88% within 2 points, Table IV) and fair reliability (ICC = 0.56, 95% CI 0.33 to 0.79) (Table V) using video-based assessment. The *Domain scores* showed good agreement (94-100% within 2 points depending on mode, Table IV). *Sensory Orientation* demonstrated excellent agreement, but low reliability, again due to a ceiling effect with most children scoring full points. *Reactive Postural Responses* also showed the lowest agreement and low reliability. The SDC for the Mini-BESTest *Total Score* was good (2.7, 10%), which indicates that children must improve by 3 points to demonstrate real change when assessed by two examiners.

3.2.3 Test-retest reproducibility (real-time and video assessment)

The Mini-BESTest *Total Score* showed slightly better test-retest agreement for real-time (100% within 2 points) than video assessment (91% within 2 points) (Table IV). Reliability was excellent for real-time (ICC = 0.84, 95% CI 0.72 to 0.96) and good for video assessment (ICC = 0.74, 95% CI 0.55 to 0.93) (Table V). The *Domain scores* showed excellent agreement for both real-time (both modes 100% within 2 points) and video assessment (91-100% within 2 points, Table IV). When assessed in real-time three domains showed good-excellent reliability, except for *Sensory Orientation*, which demonstrated a ceiling effect as noted previously. When assessed using video, one domain showed excellent reliability and the remaining three domains demonstrated poor reliability. The SDC was excellent for real-time (SDC 1.3, 5%) and good for video assessment (SDC 2.4, 9%), indicating children must improve by 2 points on the 28-point scale to demonstrate real change when assessed in real-time on different days.

4. **Discussion**

This study was the first to examine the reproducibility of the Full-BESTest and the Mini-BESTest postural control test batteries for school-aged children with typical development. Results showed that both test versions can discriminate postural control abilities in typically developing children and both have the potential to measure change over time, but the Full-BESTest has important advantages.

Our results suggest that although both BESTest versions can discriminate postural control abilities in typically developing children, the Full-BESTest (ICCs = 0.82 to

0.96) may show slightly superior reliability compared to the Mini-BESTest (ICCs = 0.56 to 0.86)). Studies in healthy adults mirror this finding with higher reliability for the Full-BESTest (ICC=0.77 to 0.86) compared to the Mini-BESTest (ICC=0.71 to 0.73) [41]. Similar results were found in adults with neurological conditions on the Full-BESTest (ICC=0.88-0.99) [42, 43],[31] compared to the Mini-BESTest (ICC range 0.72- 0.99) [44]. Based on these findings, the Full-BESTest is recommended to distinguish different stages of postural control development in typical children. Further research is recommended to establish reliability for children with motor impairments.

Both BESTest versions showed good to excellent test agreement over days, which provides a good basis for detecting changes in postural control over time. Smallest detectable changes across days were good-excellent for both the Full-BESTest (5-6%) and Mini-BESTest (5-9%) for real-time and video ratings. This data is better than that previously reported for healthy older adults using the Full-BESTest (real time assessment SDC = 8%)[41] and the Mini-BESTest (systematic review results ranging from 10.7% to 14.6%) [44]. So, both test versions may be considered appropriate to monitor postural control development in children over time. The Full-BESTest and Mini-BESTest may also be appropriate tools for measuring responsiveness to intervention, however first, pre-post intervention research in specific paediatric populations is recommended as has been conducted for adults participating in stroke rehabilitation [33].

For treatment purposes, it is important to know the reliability of the six postural control domains assessed by the Full-BESTest and the four assessed by the Mini-BESTest.

Test-retest reliability was fair to excellent for all Full-BESTest domains (ICCs >0.47), except for *Reactive Postural Responses* (ICC= 0.28 video and real-time) and *Sensory Orientation* domains (ICC= 0.30 real-time). For the *Sensory Orientation* domain, low reliability was due to a ceiling effect, where most children scored 100% on both occasions . Although this high level of sensory function is expected for children older than seven years [3], this ceiling effect may not be an issue for children with motor disorders. Adults with neurological impairments are known not to demonstrate a ceiling effect, with performance variability producing excellent reliability (ICCs >0.79) [31] [43]. The *Reactive Postural Responses* domain did not reach a ceiling effect, rather children did show more variability on this item between days. Some items in this domain may need minor modifications to improve reliability in children.

4.1 Strengths, limitations and future directions for research

The results of this study support the preferential use of the Full-BESTest over the Mini-BEStest when working with children. Ideally, scoring would be performed by video following real-time administration of the test. There may be some minor modifications needed to improve reproducibility when applying the test with children since the original test versions were designed for adults. For example, in the 6-item *Reactive Postural Response* domain, the “in-place response forward” item (number 14) and “in-place response backward” item (number 15) demonstrated the worst agreement in most test conditions. Both items required the child to hold a position against resistance and then react to the postural disturbance when released unpredictably. This was difficult for several children to perform. Re-analysis after adjusting the level of difficulty for the highest rating on these two items improved agreement from fair to excellent (Appendix 3). We propose this slightly modified version of the BESTest be termed the Kids-BESTest, suitable for children aged 8-14 years. All suggested changes to the Full-BESTest to form the Kids-BESTest have been indicated using highlighting in Appendix 3. A reformatted and print ready copy of the Kids-BESTest is contained in Appendix 4. Further research on reliability of the Kids-BESTest in a range of clinical groups is needed.

5. Conclusion

This study was the first to examine the reproducibility of the Full-BESTest and Mini-BESTest for typically developing children. Results showed that both test versions can discriminate postural control abilities in children with typical development, and are both potentially able to demonstrate change in postural control function between days. The Full-BESTest is the most comprehensive test version and it appears to show marginally better reproducibility than the Mini-BESTest, therefore it may be the optimal version when working with children. However, further psychometric research is required to confirm this with clinical paediatric populations.

Competing interests: None

Source(s) of support: We acknowledge funding support awarded by the Research Foundation, Cerebral Palsy Alliance (PG4114) through the Children's Motor Control Research Collaboration.

Acknowledgements: We would like to acknowledge the Children's Motor Control Research Collaboration and The University of Queensland Laboratory for Motor

Control and Pain Research for support to conduct the study. We would also like to thank Anne Kelly BSC (Physiotherapy) MPHTY (Paediatrics) APAM titled paediatric physiotherapist, for her assistance as the second rater for the BESTest. Most importantly, we would like to thank the children and parents who participated in this study.

SUPPORTING INFORMATION

Appendix 1. Modifications made to the Full-BESTest prior to reliability testing

Appendix 2. Full-BESTest video recording protocol

Appendix 3. Kids-BESTest highlighted modifications

Appendix 4. Kids-BESTest printable version

References

[1] Dewar R, Claus AP, Tucker K, Johnston LM. Perspectives on postural control dysfunction to inform future research: A Delphi study for children with cerebral palsy. Arch Phys Med Rehab. 2016.

[2] Shumway-Cook A, Woollacott MH. Normal Postural Control. Motor control: translating research into clinical practice. Fourth Edition ed. Baltimore: Lippincott Williams & Wilkins; 2012.

[3] Westcott SL, Burtner P. Postural control in children. Phys Occup Ther Pediatr. 2004;24:5-55.

[4] Rosenbaum P, Paneth N, Leviton A, Goldstein M, Bax M, Damiano D, et al. A report: the definition and classification of cerebral palsy April 2006. Dev Med Child Neurol Suppl. 2007;109:8-14.

[5] Johnston LM, Burns YR, Brauer SG, Richardson CA. Differences in postural control and movement performance during goal directed reaching in children with developmental coordination disorder. Hum Movement Sci. 2002;21:583-601.

[6] Kane K, Barden J. Contributions of trunk muscles to anticipatory postural control in children with and without developmental coordination disorder. Hum Mov Sci. 2012;31:707-20.

[7] Fong SS, Ng SS, Guo X, Wang Y, Chung RC, Stat G, et al. Deficits in Lower Limb Muscle Reflex Contraction Latency and Peak Force Are Associated With Impairments in Postural Control and Gross Motor Skills of Children With Developmental Coordination Disorder: A Cross-Sectional Study. Medicine (Baltimore). 2015;94:e1785.

[8] Norrlin S, Karlsson A, Ahlsten G, Lanshammar H, Silander HC, Dahl M. Force measurements of postural sway and rapid arm lift in seated children with and without MMC. Clin Biomech (Bristol, Avon). 2002;17:197-202.

[9] Shumway-Cook A, Woollacott MH. Dynamics of postural control in the child with Down syndrome. Phys Ther. 1985;65:1315-22.

[10] Freitag CM, Kleser C, Schneider M, von Gontard A. Quantitative assessment of neuromotor function in adolescents with high functioning autism and Asperger Syndrome. J Autism Dev Disord. 2007;37:948-59.

[11] Lorefice LE, Galea MP, Clark RA, Doyle LW, Anderson PJ, Spittle AJ. Postural control at 4 years in very preterm children compared with term-born peers. Dev Med Child Neurol. 2015;57:175-80.

[12] Wang TN, Howe TH, Hinojosa J, Hsu YW. Postural control of pre-term infants at 6 and 12 months corrected age. Early Hum Dev. 2010;86:433-7.

[13] Melo Rde S, Lemos A, Macky CF, Raposo MC, Ferraz KM. Postural control assessment in students with normal hearing and sensorineural hearing loss. Braz J Otorhinolaryngol. 2015;81:431-8.

[14] Bigongiari A, De Andrade E Souza F, Franciulli PM, Neto SER, Araujo RC, Mochizuki L. Anticipatory and compensatory postural adjustments in sitting in children with cerebral palsy. Hum Movement Sci. 2011;30:988-97.

[15] Girolami GL, Shiratori T, Aruin AS. Anticipatory postural adjustments in children with hemiplegia and diplegia. J Electromyogr Kinesiol. 2011;21:988-97.

[16] Tomita H, Fukaya Y, Honma S, Ueda T, Yamamoto Y, Shionoya K. Anticipatory postural muscle activity associated with bilateral arm flexion while standing in individuals with spastic diplegic cerebral palsy: A pilot study. Neurosci Lett. 2010;479:166-70.

[17] Westcott SL, Zaino CA, Miller F, Thorpe D, Unanue R. Anticipatory postural coordination and functional movement skills by degree of cerebral palsy in children age 6-12 years. Society for Neuroscience Abstracts1998. p. 149.

[18] Westcott SL, Zaino CA, Unanue R, Thorpe D, Miller F. Comparison of anticipatory postural control and dynamic balance ability in children with and without cerebral palsy. Dev Med Child Neurol. 1998;40:14.

[19] Burtner P, Qualls C, Woollacott M. Muscle activation characteristics of stance balance control in children with spastic cerebral palsy. Gait Posture. 1998;8:163-74.

[20] Burtner P, Woollacott M, Craft G, Roncesvalles M. The capacity to adapt to changing balance threats: A comparison of children with cerebral palsy and typically developing children. Dev Neurorehabil. 2007;10:249-60.

[21] Roncesvalles M, Woollacott M, Burtner P. Neural factors underlying reduced postural adaptability in children with cerebral palsy. Neuroreport. 2002;13:2407-10.

[22] Lowes LP, Westcott SL, Palisano RJ, Effgen SK, Orlin MN. Muscle force and range of motion as predictors of standing balance in children with cerebral palsy. Phys Occup Ther Pediatr. 2004;24:57-77.

[23] Lowes LP, Habib Z, Bleakney D, Westcott S. Relationship between clinical measures of balance and functional abilities in children with cerebral palsy. Pediatr Phys Ther. 1996;8:176.

[24] Westcott SL, Zaino CA, Miller F, Thorpe DE. Comparison of Postural Muscle Activity in Children of Different Ages During Stand and Reach From Firm. Compliant, and Narrow Surfaces: 9. Pediatr Phys Ther. 1997;9:207.

[25] Westcott SL, Lowes LP, Richardson PK. Evaluation of postural stability in children: current theories and assessment tools. Phys Ther. 1997;77:629-45.

[26] Saether R, Helbostad JL, Riphagen II, Vik T. Clinical tools to assess balance in children and adults with cerebral palsy: a systematic review. Dev Med Child Neurol. 2013.

[27] Geuze RH. Postural control in children with developmental coordination disorder. Neural Plast. 2005;12:183-96; discussion 263-72.

[28] Dewar R, Love S, Johnston LM. Exercise interventions improve postural control in children with cerebral palsy: a systematic review. Dev Med Child Neurol. 2015;57:504-20.

[29] Pavao SL, dos Santos AN, Woollacott MH, Rocha NACF. Assessment of postural control in children with cerebral palsy: A review. Res Dev Disabil. 2013;34:1367-75.

[30] Dewar R, Claus A, Tucker K, Johnston L. Perspectives on postural control dysfunction to inform future research: A Delphi study for children with cerebral palsy. Arch Phys Med Rehab.In press.

[31] Horak FB, Wrisley DM, Frank J. The balance evaluation systems test (BESTest) to differentiate balance deficits. Phys Ther. 2009;89:484-98.

[32] Tamura T, Otaka Y, Konno S, Sadashima K, Tomatsu T, Machida S. The Impaired Balance Systems Identified by the BESTest in Older Patients With Knee Osteoarthritis. PM R. 2016;8:869-75.

[33] Chinsongkram B, Chaikeeree N, Saengsirisuwan V, Horak FB, Boonsinsukh R. Responsiveness of the Balance Evaluation Systems Test (BESTest) in People With Subacute Stroke. Phys Ther. 2016.

[34] Kurz MJ, Corr B, Stuberg W, Volkman KG, Smith N. Evaluation of lower body positive pressure supported treadmill training for children with cerebral palsy. Pediatr Phys Ther. 2011;23:232-9.

[35] Franchignoni F, Horak F, Godi M, Nardone A, Giordano A. Using psychometric techniques to improve the Balance Evaluation System Test: the mini-BESTest. J Rehabil Med. 2010;42:323.

[36] Padgett PK, Jacobs JV, Kasser SL. Is the BESTest at its best? A suggested brief version based on interrater reliability, validity, internal consistency, and theoretical construct. Phys Ther. 2012;92:1197-207.

[37] de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol. 2006;59:1033-9.

[38] Horak FB.

[39] Cicchetti DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. Am J Ment Defic. 1981;86:127-37.

[40] Auld ML, Ware RS, Boyd RN, Moseley GL, Johnston LM. Reproducibility of tactile assessments for children with unilateral cerebral palsy. Phys Occup Ther Pediatr. 2012;32:151-66.

[41] Marques A, Almeida S, Carvalho J, Cruz J, Oliveira A, Jácome C. Balance tests in healthy older people Reliability, validity and ability to identify fall status of the BESTest, Mini-BESTest and Brief-BESTest in older people living in the community. Arch Phys Med Rehab. 2016.

[42] Leddy AL, Crowner BE, Earhart GM. Functional gait assessment and balance evaluation system test: reliability, validity, sensitivity, and specificity for identifying individuals with Parkinson disease who fall. Phys Ther. 2011;91:102-13.

[43] Chinsongkram B, Chaikeeree N, Saengsirisuwan V, Viriyatharakij N, Horak FB, Boonsinsukh R. Reliability and validity of the Balance Evaluation Systems Test (BESTest) in people with subacute stroke. Phys Ther. 2014;94:1632-43.

[44] Di Carlo S, Bravini E, Vercelli S, Massazza G, Ferriero G. The Mini-BESTest: a review of psychometric properties. International Journal of Rehabilitation Research. 2016;39:97-105.

[45] Williams EN, Carroll SG, Reddihough DS, Phillips BA, Galea MP. Investigation of the timed 'Up & Go' test in children. Dev Med Child Neurol. 2005;47:518-24.

Table I. Summary of participant characteristics

	Day 1 Intra-rater video Inter-rater video	Day 2 Test-retest real-time Test-retest video
(n)	34	22
Male, n(%)	19 (56%)	10 (45%)
Age, mean (SD)	10.8 (2.2) years	10.9 (1.6) years
Body mass index, mean (SD)	16.4 (3.7)	16.7 (4.2)
Height, mean (SD)	149.2 (13.4) cm	149.0 (13.3) cm
Weight, mean (SD)	37.3 (14) kg	38.1 (16) kg
BOT-2* % rank, mean (SD)	52 (25) percentile	52 (24) percentile

BOT-2, Bruininks-Oseretsky Test of Motor Proficiency, second Edition

Table II. Agreement of the Full-BESTest

Full-BESTest	Increment (Range)	Agreement (%)			SEM	SDC(%)*		95% CI for LOA
		Exact	1 point	2 points				
1. Intra-rater agreement (n= 34, video 1, one assessor)								
Biomechanical Constraints	1 (0-15)	85	97	100	0.35	1.0	(7%)	-1.0 to 1.1
Stability Limits and Verticality	1 (0-21)	91	97	100	0.29	0.8	(4%)	-0.7 to 1.0
Transitions/Anticipatory	1 (0-18)	50	91	100	0.63	1.8	(10%)	-1.9 to 1.9
Reactive	1 (0-18)	74	97	100	0.41	1.1	(6%)	-1.0 to 1.4
Sensory Orientation	1 (0-15)	88	100	100	0.24	0.7	(5%)	-0.8 to 0.7
Stability in Gait	1 (0-21)	62	94	100	0.52	1.4	(7%)	-1.7 to 1.3
Total score	1 (0-108)	20	79	100	0.81	2.2	(2%)	-2.3 to 2.5
2. Inter-rater agreement (n=34, video 1, two assessors)								
Biomechanical Constraints	1 (0-15)	56	85	100	0.67	1.9	(13%)	-2.1 to 1.9
Stability Limits and Verticality	1 (0-21)	74	79	100	0.66	1.8	(9%)	-1.8 to 2.1
Transitions/Anticipatory	1 (0-18)	53	97	97	0.58	1.6	(9%)	-2.0 to 1.5
Reactive	1 (0-18)	50	82	88	0.85	2.3	(13%)	-1.9 to 3.2
Sensory Orientation	1 (0-15)	62	97	100	0.42	1.2	(8%)	-1.6 to 0.9
Stability in Gait	1 (0-21)	41	76	94	0.94	2.6	(12%)	-3.0 to 2.5
Total score	1 (0-108)	24	50	82#	1.45	4.0	(4%)	-4.4 to 4.1
3. Test-retest agreement (n=22, video 1 and 2, one assessor)								
Biomechanical Constraints	1 (0-15)	64	95	95	0.62	1.7	(12%)	-1.7 to 1.9
Stability Limits and Verticality	1 (0-21)	95	95	100	0.30	0.8	(4%)	-1.0 to 0.8
Transitions/Anticipatory	1 (0-18)	36	73	100	0.87	2.4	(13%)	-2.5 to 2.7
Reactive	1 (0-18)	27	68	91	1.44	4.0	(22%)	-4.2 to 4.3
Sensory Orientation	1 (0-15)	59	100	100	0.45	1.3	(9%)	-1.2 to 1.2
Stability in Gait	1 (0-21)	63	95	100	0.83	2.3	(11%)	-2.6 to 2.3
Total score	1 (0-108)	14	36	64#	1.77	4.9	(5%)	-5.4 to 5.1
4. Test-retest agreement (n= 22, real-time, one assessor)								
Biomechanical Constraints	1 (0-15)	81	100	100	0.30	0.8	(5%)	-1.0 to 0.8
Stability Limits and Verticality	1 (0-21)	54	77	100	0.77	2.1	(10%)	-2.2 to 2.1
Transitions/Anticipatory	1 (0-18)	54	77	100	0.75	2.1	(12%)	-1.9 to 2.5
Reactive	1 (0-18)	18	59	86	1.63	4.5	(25%)	-4.3 to 5.3
Sensory Orientation	1 (0-15)	59	91	100	0.59	1.6	(11%)	-1.9 to 1.6
Stability in Gait	1 (0-21)	36	68	95	0.99	2.8	(13%)	-2.4 to 3.4
Total score	1 (0-108)	18	36	54	2.38	6.6	(6%)	-5.6 to 8.1

SEM, standard error of the mean; SDC, smallest detectable change; CI, confidence interval; LOA, limits of agreement; SDC is expressed as a percentage of the Total score or domain score to allow comparison of scores with different ranges;# 90% or more of scores agreed within 4 points.*

Table III. Reliability of the Full-BESTest

Full-BESTest	ICC	95% CI
1. Intra-rater reliability (n=34, video 1, one assessor)		
Biomechanical Constraints	0.84	0.74 to 0.99
Stability Limits and Verticality	0.88	0.81 to 0.96
Transitions/Anticipatory	0.80	0.67 to 0.92
Reactive	0.90	0.83 t0 0.96
Sensory Orientation	0.84	0.74 to 0.94
Stability in Gait	0.82	0.70 to 0.93
Total score	0.96	0.93 to 0.99
2. Inter-rater reliability (n= 34, video 1, two assessor)		
Biomechanical Constraints	0.30	0.00 to 0.61
Stability Limits and Verticality	0.52	0.27 to 0.76
Transitions/Anticipatory	0.83	0.72 to 0.94
Reactive	0.65	0.45 to 0.84
Sensory Orientation	0.22	0.00 to 0.54
Stability in Gait	0.66	0.39 to 0.82
Total score	0.87	0.79 to 0.95
3. Test-retest reliability (n=22, video 1 and 2, one assessor)		
Biomechanical Constraints	0.61	0.35 to 0.88
Stability Limits and Verticality	0.80	0.65 to 0.95
Transitions/Anticipatory	0.68	0.46 to 0.91
Reactive	0.28	0.00 to 0.67
Sensory Orientation	0.47	0.14 to 0.80
Stability in Gait	0.72	0.52 to 0.93
Total score	0.84	0.72 to 0.96
4. Test-retest reliability (n= 22, real-time, one assessor)		
Biomechanical Constraints	0.82	0.68 to 0.96
Stability Limits and Verticality	0.45	0.12 to 0.79
Transitions/Anticipatory	0.78	0.62 to 0.95
Reactive	0.34	0.00 to 0.72
Sensory Orientation	0.30	0.00 to 0.68
Stability in Gait	0.76	0.58 to 0.94
Total score	0.82	0.69 to 0.96

ICC, Intra-class correlation coefficient; CI, Confidence Interval.

Table IV. Agreement of the Mini-BESTest

Mini-BESTest	Increment (Range)	Agreement (%)			SEM	SDC (%)*	95% CI of LOA (95% CI)
		Exact	1 point	2 points			
1. Intra-rater agreement (n=34, video 1, one assessor)							
Transitions/Anticipatory	1 (0-6)	74	100	100	0.37	1.0 (17%)	-1.1 to 1.1
Reactive	1 (0-6)	82	100	100	0.29	0.8 (13%)	-0.7 to 1.0
Sensory Orientation	1 (0-6)	100	100	100	0	0	0
Stability in Gait	1 (0-10)	71	97	100	0.45	1.3 (13%)	-1.3 to 1.3
Total score	1 (0-28)	50	97	100	0.54	1.5 (5%)	-1.4 to 1.8
2. Inter-rater agreement (n=34, video 1, two assessors)							
Transitions/Anticipatory	1 (0-6)	71	100	100	0.38	1.1 (18%)	-1.0 to 1.2
Reactive	1 (0-6)	62	85	94	0.63	1.7 (29%)	-1.3 to 2.4
Sensory Orientation	1 (0-6)	94	100	100	0.17	0.5 (8%)	-0.4 to 0.6
Stability in Gait	1 (0-10)	56	91	100	0.60	1.7 (17%)	-1.8 to 1.8
Total score	1 (0-28)	35	68	88	0.96	2.7 (10%)	-2.1 to 3.6
3. Test-retest agreement (n=22, video 1 and 2, one assessor)							
Transitions/Anticipatory	1 (0-6)	55	91	95	0.62	1.7 (29%)	-1.7 to 1.9
Reactive	1 (0-6)	55	95	100	0.55	1.5 (25%)	-1.7 to 1.6
Sensory Orientation	1 (0-6)	100	100	100	0	0	0
Stability in Gait	1 (0-10)	63	95	100	0.50	1.4 (14%)	-1.6 to 1.3
Total score	1 (0-28)	32	91	91	0.86	2.4 (9%)	-2.7 to 2.4
4. Test-retest agreement (n=22, real-time, one assessor)							
Transitions/Anticipatory	1 (0-6)	59	100	100	0.45	1.3 (21%)	-1.5 to 1.2
Reactive	1 (0-6)	68	95	100	0.47	1.3 (22%)	-1.6 to 1.2
Sensory Orientation	1 (0-6)	95	100	100	0.15	0.4 (7%)	-0.5 to 0.4
Stability in Gait	1 (0-10)	41	95	100	0.67	1.9 (19%)	-1.7 to 2.3
Total score	1 (0-28)	23	63	100	0.47	1.3 (5%)	-3.0 to 2.9

*SEM: standard error of the mean, SDC: smallest detectable change, CI: confidence interval, LOA: limits of agreement, * SDC is expressed as a percentage of the Total score or domain score to allow comparison of scores with different ranges*

Table V. Reliability of the Mini-BESTest

Mini-BESTest	ICC	95% CI
1. Intra-rater reliability (n=34, video 1, one assessor)		
Transitions/Anticipatory	0.67	0.48 to 0.86
Reactive	0.79	0.66 to 0.92
Sensory Orientation	-	-
Stability in Gait	0.80	0.67 to 0.92
Total score	0.86	0.78 to 0.95
2. Inter-rater reliability (n=34, video 1, two assessor)		
Transitions/Anticipatory	0.64	0.44 to 0.84
Reactive	0.26	0.00 to 0.58
Sensory Orientation	0.00	0.00 to 0.34
Stability in Gait	0.66	0.47 to 0.85
Total score	0.56	0.33 to 0.79
3. Test-retest reliability (n=22, video 1 and 2, one assessor)		
Transitions/Anticipatory	0.23	0.00 to 0.63
Reactive	0.12	0.00 to 0.54
Sensory Orientation	-	-
Stability in Gait	0.81	0.67 to 0.96
Total score	0.74	0.55 to 0.93
4. Test-retest reliability (n=22, real-time, one assessor)		
Transitions/Anticipatory	0.60	0.32 to 0.87
Reactive	0.61	0.35 to 0.88
Sensory Orientation	0.00	0.00 to 0.42
Stability in Gait	0.80	0.64 to 0.95
Total score	0.84	0.72 to 0.96

ICC, Intra-class correlation coefficient; CI, Confidence Interval.